REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
20-06-2003	Technical	Jun 1999-Jun 2003

4. TITLE AND SUBTITLE
Science and Technology Text Mining:
Cross-Disciplinary Innovation

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Kostoff, Ronald, N.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Unlimited Distribution

20030618 156

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Innovation is critical for maintaining competitive advantage in a high tech global economy, especially for organizations or nations that do not possess low cost labor forces. Many studies on innovation attempt to identify endogenous and exogenous variables that impact innovation (Kostoff, 1997a), in order to better understand the environment that promotes innovation. The author's recent efforts have focused on developing processes for enhancing innovation that exploit the transference of information and insights among seemingly disparate disciplines.

The objective of this report is to describe how innovation can be promoted through the enhancement of discovery by cross-discipline knowledge transfer. The approach developed entails two complementary components – one literature based, the other workshop-based. The literature-based component identifies the science and technology disciplines related to the central theme of interest, the experts in these disciplines, and promising candidate concepts for innovative solutions. These outputs define the agenda and participants for the workshop-based component. An example of this combined approach is presented for the theme of Autonomous Flying Systems. The hybrid approach appears to be an excellent vehicle for generating discovery and enabling innovation. However, it requires substantial time and effort in both phases.

15. SUBJECT TERMS

Innovation, text mining, literature-based discovery, clustering, workshops, cross-discipline, multi-disciplinary, interdisciplinary, discovery, creativity, brainstorming, network-centric, science and technology, technical literature, Database Tomography, computational linguistics, literature survey, information retrieval

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Ronald N. Kostoff
U	U	U			19b. TELEPHONE NUMBER *(include area code)* 703-696-4198

SCIENCE AND TECHNOLOGY TEXT MINING: CROSS-DISCIPLINARY INNOVATION

BY

DR. RONALD N. KOSTOFF
OFFICE OF NAVAL RESEARCH
ARLINGTON, VA 22217
PHONE: 703-696-4198
FAX: 703-696-4274
INTERNET: kostofr@onr.navy.mil

KEYWORDS: Innovation, text mining, literature-based discovery, clustering, workshops, cross-discipline, multi-disciplinary, interdisciplinary, discovery, creativity, brainstorming, network-centric, science and technology, technical literature, Database Tomography, computational linguistics, literature survey, information retrieval

ABSTRACT

Innovation is critical for maintaining competitive advantage in a high tech global economy, especially for organizations or nations that do not possess low cost labor forces. Many studies on innovation attempt to identify endogenous and exogenous variables that impact innovation (Kostoff, 1997a), in order to better understand the environment that promotes innovation. The author's recent efforts have focused on developing processes for enhancing innovation that exploit the transference of information and insights among seemingly disparate disciplines.

The objective of this report is to describe how innovation can be promoted through the enhancement of discovery by cross-discipline knowledge transfer. The approach developed entails two complementary components – one literature based, the other workshop-based. The literature-based component identifies the science and technology disciplines related to the central theme of interest, the experts in these disciplines, and promising candidate concepts for innovative solutions. These outputs define the agenda and participants for the workshop-based component. An example of this combined approach is presented for the theme of Autonomous Flying Systems. The hybrid approach appears to be an excellent vehicle for generating discovery and enabling innovation. However, it requires substantial time and effort in both phases.

INTRODUCTION

Innovation reflects the metamorphosis from present practice to some new, hopefully "better" practice. It can be based on:

1) existing non-implemented knowledge;
2) discovery of previously unknown information;
3) discovery and synthesis of publicly available knowledge whose independent segments have never been combined; and/ or
4) invention.

In turn, the invention could derive from logical exploitation of a knowledge base, and/ or from spontaneous creativity (e.g., Edisonian discoveries from trial and error).

The process of innovation is of immense social interest and impact. Classical studies by Mansfield (1980, 1991), Griliches (1958, 1979, 1994), and Terleckyj (1977, 1985) focused on the relationship between innovation and micro or macro economics. Studies by Wenger (1999) on combined visualization/ brainstorming techniques, Patton (2002) and Taggar (2001) on the impact of group stimulation to creativity, Chen (1998) and Siau (1996) on contributions of electronic technology to creativity, and books by Boden (1991) and DeBono (1992) on mental processes in creativity, focused on the process of creativity and its contributions to innovation. Large-scale studies by the Department of Defense (DoD, 1969), Illinois Institute of Technology Research Institute (IITRI, 1968), Battelle (Battelle, 1973), and the Institute for Defense Analysis (IDA, 1990, 1991a, 1991b) focused on identifying the environmental and management conditions most conducive to innovation. Recent symposia have focused on the relation of innovation to: technology policy (Conceicao, 1998, 2001); technology forecasting (Grupp and Linstone, 1999; Arciszewski, 2000), competitive advantage (Hitt et al, 2000); and economic growth and impact (Van de Klundert et al, 1998; Spender and Grant, 1996; Archibugi and Michie, 1995). Yet both the process and impacts of innovation remain poorly understood.

One of the least studied components of innovation is the discovery and synthesis of publicly available knowledge whose independent segments have never been combined; i.e., the transfer of information and understanding developed in one or more disciplines to other, perhaps very disparate, disciplines. With the explosion in availability of information, the number of opportunities to synthesize knowledge and enhance discovery from disparate disciplines increases non-linearly. Conversely, with accelerating production of information, scientists and

technologists find it increasingly difficult to remain aware of advances within their own discipline(s), much less advances in other seemingly unrelated ones. Paradoxically, the growth in science has led to the balkanization of science!

As science and technology become more specialized, the incentives for interdisciplinary research and development are reduced, and this cross-discipline transfer of information becomes more difficult. The author's observation, from examination of many science and technology sponsoring agencies and performing organizations, supplemented by a wide body of literature (Metzger, 1999; Naiman, 1999; Bauer, 1990; Bruhn, 1995; Butler, 1998), is that *strong cross-disciplinary dis-incentives exist at all phases of program/ project evolution, including selection, management and execution, review, and publication* (See Appendix 3 for a more thorough discussion of roadblocks to interdisciplinary research). To overcome cross-discipline transmission barriers, and thereby enhance innovation, systematic methods are required to heighten awareness of experts in one discipline to advances in other disciplines. Most desirable are methods that incorporate/ require cross-disciplinary access as an organic component.

This report presents two different, yet complementary, approaches to increase cross-discipline knowledge transfer and provide the framework for enhancing innovation. One is literature-based, the other is workshop-based. Each approach individually represents a major advance in enabling discovery and subsequent innovation, and the hybrid of the two approaches provides a synergy that multiplies their combined benefits.

The literature-based approach is summarized first, followed by the workshop-based approach. The advantages of combining the two approaches are then presented. The details of each approach are presented in the first two appendices. The third appendix addresses some of the intrinsic roadblocks to performing interdisciplinary research. This appendix outlines a process for determining the relationship of the disciplines required for innovation to disciplines selected for an innovation-based program of research.

ACCESSING LINKED LITERATURES FOR ENHANCING INNOVATION-SUMMARY

The first approach searches for relationships between linked, overlapping literatures, and discovers relationships or promising opportunities not obtainable from reading each literature separately. The general theory behind this approach,

applied to two separate literatures, is based upon the following considerations (Swanson, 1986).

Assume that two literatures with disjoint components can be generated, the first literature AB having a central theme "a" and sub-themes "b," and the second literature BC having a central theme(s) "b" and sub-themes "c." From these combinations, linkages can be generated through the "b" themes that connect both literatures (e.g., AB-->BC). Those linkages that connect the disjoint components of the two literatures (i.e., the components of AB and BC whose intersection is zero) are candidates for discovery, since the disjoint themes "c" identified in literature BC could not have been obtained from reading literature AB alone.

Some initial applications of the first approach have been published in the medical literature (Swanson, 1986). One interesting discovery was that dietary eicosapentaenoic acid (theme "a" from literature AB) can decrease blood viscosity (theme "b" from both literatures AB and literatures BC) and alleviate symptoms of Raynaud's disease (theme "c" from literature BC). There was no mention of eicosapentaenoic acid in the Raynaud's disease literature, but the acid was linked to the disease through the blood viscosity themes in both literatures. Subsequent medical experiments confirmed the validity of this literature-based discovery (Gordon and Lindsay, 1996). (A web site (http://kiwi.uchicago.edu/) overviews the process used to generate this discovery, and contains software that allows the user to experiment with the technique. Finn (1998) outlines perceptions of different knowledgeable individuals on Swanson and Smalheiser's general technique.)

This literature-based discovery approach is in its infancy. Public and private financial support for this technology are minimal. It is a research area of unlimited potential that seems to have fallen through the cracks. There is essentially one group that is publishing results of literature-based innovation and discovery in the credible peer-reviewed literature (Swanson, 1986, 1997, 1999; Smalheiser, 1994, 1998a, 1998b), two groups that have published concept papers (Hearst, 1999; Kostoff, 1999a), and a few other groups that have replicated Swanson's initial results (Gordon and Lindsay, 1996; Weeber et al, 2001). Presently, the approach is not automatic. It requires much thought, expertise, and effort. The author's group is examining different approaches to make the process more systematic, while reducing the manual labor intensity. Given the potential benefits of the literature-based approach for stimulating innovation, it is truly a technology whose time has come.

Appendix 1 generalizes and expands upon the literature-based approach, using the Database Tomography techniques and experience developed by the author since 1991 (Kostoff, 1993, 1994, 1998, 1999b, 2000a, 2000b). It outlines the theory of the expanded approach, the implementation details, and overviews the range of applications possible with this technique.

INTERDISCIPLINARY WORKSHOPS FOR ENHANCING INNOVATION-SUMMARY

The second approach consists of convening workshop(s) of experts from different disciplines focused on specific central themes. The purpose of such a workshop is to achieve multi-discipline synergies and cross-discipline transfers to generate promising research directions for these central themes. The theory behind this approach is described in Appendix 2. To test this theory, a workshop on Autonomous Flying Systems was convened in December 1997. Its implementation mechanics and results are described in detail in Appendix 2.

The total workshop process consisted of three phases:

(1) A two month pre-meeting e-mail phase in which each participant provided descriptions of advanced capabilities and promising research opportunities from his/her discipline to all other participants;

(2) A two-day meeting at the Office of Naval Research during which the promising opportunities identified beforehand were discussed, crystallized, and enhanced; and

3) A post meeting e-mail phase in which each participant provided additional or embellished opportunities.

A number of important lessons were extracted from the conduct of this workshop, and they can be summarized as follows:

a) The workshop approach broke new ground toward stimulating innovative thought. The combination of a common theme that underlay many diverse disciplines with the guided pre-meeting cross-fertilization of ideas among these disciplines, enhanced by the intensive real-time exchange of ideas at the workshop, provided an environment highly conducive to innovation. It was not easy, simple, or effortless, and required substantial planning and work in order to be effective. One should not throw people from fifteen different disciplines

together in a room for two days and hope to get new ideas synthesized, as some modern brainstorming approaches attempt to do. There needs to be a common generic thread woven through the different disciplines represented to spark the innovative thought process.

Interdisciplinary workshops, when performed correctly, are the wave of the future in defining new research (and technology) areas and approaches. Because of the intensity and effort involved throughout the process, they are most appropriate for large scale "grand challenges" in full-blown workshop form, but are appropriate as well for smaller scale issues.

b) Representatives from diverse technical disciplines, organizations, and development categories attended the workshop. There was substantial value in having a balance of discipline, category, and organization diversity at the same meeting. The different perspectives presented benefited all participants.

The use of modern information technology can expand the degree of diversity dramatically. Some of the concepts and group software proposed for network-centric peer review (Kostoff et al, 2001c) can be easily adapted for use in innovation workshops. This would allow many more people, disciplines, and organizations to be represented, further enhancing the potential for cross-discipline information transfer and resultant innovation and discovery. Having a network-centric pre-meeting phase in tandem with a network-centric workshop would further guarantee that the interactions would be documented, including the time sequencing of its generation. This information could be analyzed further after the workshop to extract additional insights.

c) Problem selection is crucial. The problem should be sufficiently general that many diverse disciplines can link to it. Given the choice of equally relevant problems, there is more potential for impact in selecting problem areas for which a large interdisciplinary community is not yet obvious.

d) It is important to select participants by the most objective processes available. A combination of expert recommendation and strategic topical maps based on computational linguistics, publications, and citations was used for the selection process, and this approach produced highly knowledgeable individuals. Incorporation of the full literature-based approach to innovation in the discipline or participant selection process could further enhance confidence that the most appropriate mix of disciplines and experts has been chosen.

e) It is extremely important that individuals selected for participation be world-class experts in their particular areas. There are relatively very few individuals producing the seminal works in any field (Kostoff, 1998, 1999b, 2000a, 2002a), and it is these people who should be central to any truly innovative workshops. However, in addition to these established experts, highly competent individuals new to the field should also be selected. One benefit of transcending selection of known experts is that fresh faces new to established communities appear. They can sometimes challenge established paradigms and offer concepts typically not advanced through panels based solely upon well-known, over-used panelists.

f) The e-mail component of the workshop is crucial. The gestation period between the input of promising ideas and their actual discussion at the workshop allows consideration of many different approaches and syntheses. It also saves substantial time at the workshop by clarifying confusing issues beforehand. However, in the first experience reported here, the stimulation of dialogue in the e-mail phase among most of the participants did not occur. The only participant to raise questions was the author, and this occurred only a few times. Nonetheless, in these instances, the dialogue was extremely valuable in clarifying issues and surfacing points of contention. In future workshops, it is strongly recommended that a few individuals representing different disciplines be asked to assume a role of facilitator, with the task of stimulating dialogue and raising questions during the workshop build-up phase.

g) All the attendees at the workshop were required to participate; there were no pure observers. This meant that they had to submit accomplishments and opportunities statements by e-mail. They also had to be prepared to lead discussions at the workshop. This participation requirement was valuable in that each attendee obtained a sense of ownership in the workshop and its outcome. His/her contribution tended to be more substantive and creative than is typically the case at standard workshops. Those who contributed more in the e-mail phase tended to contribute more in the workshop phase. In addition, there was a sense of equality among participants when all were required to contribute, as opposed to an audience/performer environment with passive onlookers. The requirement that each attendee be an active participant translates directly into a limitation on audience size. However, it was concluded that the participation of a limited number of motivated and active individuals contributed more to the innovation process than the standard workshop of few active participants and many observers. Having network-centric operation, and the inclusion of larger numbers of people, would not contradict the requirement for active participation. Network-centric

operation including group software allows parallel inputs of information from many participants.

h) In general, there needs to be some incentive to motivate participation of world-class experts in these workshops. Unless they are able to envision some type of substantive impact resulting from their participation, either on larger science and technology issues or in their individual disciplines, they could be reluctant to invest the substantial amount of time required for serious participation. This, however, did not turn out to be a problem for the Autonomous Flying Systems workshop, apparently because of the limited size of the field and the interest of the participants in the type of workshop conducted.

In addition, during the workshop, participants did not appear to have reluctance in sharing new concepts. This is in stark contrast to some workshops the author has attended where novel ideas were held very closely. In the Autonomous Flying Systems workshop, there was a spirit of cameraderie and cooperation that pervaded the proceedings, and helped overcome the barriers to sharing. This spirit was fostered in the pre-meeting e-mail dialogue phase, and further nurtured during the meeting by having all attendees participate in the proceedings as equal partners.

Finally, interdisciplinary workshops are a powerful potential source of radically innovative ideas if conducted properly. There are three central requirements for success:

(1) A problem of significant interest to the sponsoring organization must be selected;
(2) An optimal mix of world-class experts appropriate to the problem must be chosen;
(3) Conditions must be created that will motivate the participants to share their novel concepts.

The Autonomous Flying Systems workshop addressed these three requirements to a significant degree. A preliminary concept proposal emerged, and a copy of this proposal is available from the author.

NEED FOR LITERATURE/WORKSHOP SYNERGY

Most organizations use some variant of a workshop/group dynamics approach for brain-storming or other proxies for stimulating innovation. The most current

information is available, and real-time information exchange is unmatched. The attendees and participants in these groups tend to be focused subject experts representing a small fraction of the relevant technical community; there is rarely any complementary sophisticated literature analysis performed, and there are rarely experts present from strongly divergent disciplines. The outputs and discussion are highly subjective. The workshop techniques tend not to make full use of many of the information technology advances of recent years. Probably most importantly, there are strong disincentives for the participants to reveal the latest innovations. What many workshops produce in practice are forums for "selling" completed or near-completed research efforts.

A few performers, individuals or small groups of individuals, pursue the literature-based computer-assisted approach. This literature approach tends to be more sophisticated and technologically advanced than the workshop approach, and is more objective. It is more comprehensive, since it encompasses science and technology beyond the scope of any individual, or group of individuals, and can access data from many technical disciplines and many global sources. The source data is not as current as the workshop approach, due to the documentation time lag. However, with the advent of extensive on-line documentation, this time lag has been reduced considerably. One intrinsic limitation is that only a relatively modest amount of science and technology performed globally is documented and readily accessible to the wider user community (Kostoff, 2000c); obviously, any science and technology not documented cannot be accessed. The literature-based approach has not received widespread attention and may fall short of the interpretive and analytical strengths of the workshop approach. As a result, the literature approach is not widely used (e.g., Finn, 1998).

While either the workshop approach or the literature approach can be done independently to help stimulate discovery, they should be done in tandem to maximize the benefit provided by each. There is nothing on record to indicate that this joint approach to innovation has been implemented, or even considered. The Autonomous Flying Systems workshop described in this chapter has some elements of the combined approach. Some of the Database Tomography proximity analysis tools were used to identify the scope of related literatures, and the prolific individuals in these literatures. These individuals were then invited to the workshop. However, time constraints precluded using the full capabilities that the literature-based approach can offer.

In a joint workshop-literature effort, the literature approach would be included in the background pre-meeting phase of the workshop approach (as developed in Appendix 2). Accordingly, the literature study would provide:

(1) Background reading for the workshop participants in related yet disparate science and technology areas;

(2) Strategic maps of the broader science and technology literature as outlined in the DT papers referenced above;

(3) Promising opportunities for innovation and discovery; and

(4) The disparate science and technology disciplines from which the experts for the workshop could be drawn.

The hybrid literature-workshop approach would eliminate the limitations of each approach done separately. The right people from the right combination of disciplines could be identified by the literature-based approach, and invited to the workshop. The literature-based analysis could structure the technical relationships, and provide an objective starting point for discussion. Network-centric peer review would allow linking, and fusing information from, large numbers of reviewers to incorporate more representative opinion sampling from the larger technical community. The only limitation not overcome is the disincentive for the participants, or document authors, to reveal their latest science and technology advancements.

There is extra time and cost involved with two approaches, and if responses were required with severe time limitations, then only one approach might prove feasible. For organizations that are serious about stimulating discovery and subsequent innovation, the additional time should not be a factor, given the potential high marginal benefits. Government could probably draw upon a more eclectic group than industry. Because of the competitive aspects, industry would probably rely more upon internal participants and contracted consultants, whereas government would draw upon individuals from many organizations.

CONCLUSIONS

The advent of large databases, and the parallel advances in computer hardware and software, provide the opportunity to augment and amplify traditional approaches of human creativity in generating discovery and subsequent innovation. This chapter has shown that multi-discipline structured workshops can enhance the

science and technology discovery and subsequent innovation processes, and has shown that multi-discipline literature-based analyses can enhance the science and technology discovery process. The document has shown conceptually that the combination of computer-enhanced literature-based analyses and multi-discipline structured workshops has the synergistic potential to dramatically improve the discovery and subsequent innovation process relative to the already strong capabilities available from each process separately. This literature-workshop synergy represents a potential major breakthrough for systematically identifying: 1) the most promising disciplines to be used in the workshop; 2) specific experts from these different disciplines; 3) candidate promising concepts that form the basis for discussion.

(The views expressed in this report are those of the author and do not represent the views of the Department of the Navy.)

BIBLIOGRAPHY

Archibugi, D., Michie, J. (1995). (Eds.). *Technology and innovation.* Special Issue, Cambridge Journal Of Economics. 19(1) 1-4 February.

Arciszewski, T. (2000). (Ed.) *Innovation: The key to progress in technology and society.* Special issue, Technological Forecasting And Social Change. 64: (2-3). 119-120. June-July.

Battelle (1973). *Interactions of science and technology in the innovative process: some case studies.* Final Report. Prepared for the National Science Foundation. Contract NSF-C 667. Battelle Columbus Laboratories. March 19.

Bauer, H. H. (1990). Barriers against interdisciplinarity - implications for studies of science, technology, and society (sts). *Science, Technology, and Human Values.* 15(1). Winter. 105-119.

Boden, M. (1991). *The creative mind.* New York: Basic Books.

Bruhn, J. G. (1995). Beyond discipline: Creating a culture for interdisciplinary research. *Integrative Physiological and Behavioral Science.* 30(4). September-December. 331-341.

Butler, D. 1998. Interdisciplinary research 'being stifled'. *Nature.* 396. 19 November. 202.

Chen, Z. (1998). Toward a better understanding of idea processors. *Information And Software Technology*. 40(10) 541-553. October 15.

Collins JR. (2002). May you live in interesting times: using multidisciplinary and interdisciplinary programs to cope with change in the life sciences. *BioScience*. 52:1. 75-83.

Conceicao, P., Heitor, M.V., Gibson, D.V., Shariq, S.S. (1998). (Eds.). *The emerging importance of knowledge for development: Implications for technology policy and innovation*. Special Issue, Technological Forecasting And Social Change. 58: (3). 181-202. July.

Conceicao, P., Gibson D.V., Heitor, M.V., Sirilli, G. (2001). (Eds.) *Beyond the digital economy: A perspective on innovation for the learning society*. Special Issue, Technological Forecasting And Social Change. 67(2-3). 115-142. June-July.

DeBono, E. (1992). *Serious Creativity*. New York: Harper Collins.

DOD (1969). *Project Hindsight*. Office of the Director of Defense Research and Engineering. Wash., D. C. DTIC No. AD495905. October.

Finn, R. (1998). Program uncovers hidden connections in the literature. *The Scientist*. 11 May.

Gordon, M. D., and R. K. Lindsay, (1996). Toward discovery support systems: a replication, re-examination, and extension of Swanson's work on literature-based discovery of a connection between Raynaud's disease and fish oil. *Journal of the American Society for Information Science*. 47(2). 116-128.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *The Bell Journal of Economics*. 10. Spring.

Griliches, Z. (1994). Productivity, R&D, and the data constraint. *The American Economic Review*. 84(1). March.

Griliches, Z. (1958). Research costs and social returns: hybrid corn and related innovations. *Journal of Political Economy*. 66.

Grupp, H., Linstone, H.A. (1999). (Eds.) National technology foresight activities around the globe - Resurrection and new paradigms. Special Issue, *Technological Forecasting And Social Change*. 60(1). 85-94. January.

Hearst, M. A., (1999). Untangling Text Data Mining. *Proceedings of ACL 99, the 37th Annual Meeting of the Association for Computational Linguistics*. University of Maryland. June 20-26. 1-9.

Hitt, M.A., Ireland, R.D., Lee, H.U. (2000). (Eds.). *Technological learning, knowledge management, firm growth and performance*. Special issue, Journal Of Engineering And Technology Management. 17(3-4). 231-246. September-December.

IDA. *DARPA technical accomplishments*. Volume I. IDA Paper P-2192. February 1990; Volume II. IDA Paper P-2429. April 1991; Volume III. IDA Paper P-2538. July 1991. Institute for Defense Analysis.

IITRI (1968). *Technology in retrospect and critical events in science*, Illinois Institute of Technology Research Institute Report. December.

Kostoff, R. N. (1993). Database tomography for technical intelligence. *Competitive Intelligence Review*. 4(1).

Kostoff, R.N. (1994). Database tomography: origins and applications. *Competitive Intelligence Review*. Special Issue on Technology. 5(1).

Kostoff, R. N., (1997a). *The handbook of research impact assessment*. Seventh Edition. DTIC Report Number ADA296021. Summer. Also located at www.dtic.mil/dtic/kostoff/index.html.

Kostoff, R. N. (1997b). Database tomography for information retrieval. *Journal of Information Science*. 23(4).

Kostoff RN. (1997). Peer review: the appropriate GPRA metric for research. *Science*. 277. 1 August. 651-652.

Kostoff, R. N. (1998). Database tomography for technical intelligence: a roadmap of the near-earth space science and technology literature. *Information Processing and Management*. 34(1).

Kostoff, R. N. (1999a). Science and technology innovation. *Technovation*. 19:10. October. 593-604.

Kostoff, R. N. (1999b). Hypersonic and supersonic flow roadmaps using bibliometrics and database tomography. *Journal of the American Society for Information Science*. 50(5). April.

Kostoff, R. N., Braun, T., Schubert, A., Toothman, D. R., and Humenik, J. (2000a). Fullerene roadmaps using bibliometrics and database tomography. *Journal of Chemical Information and Computer Science*. 40. January-February.

Kostoff, R. N., Green, K. A., Toothman, D. R., and Humenik, J. (2000b). Database tomography applied to an aircraft science and technology investment strategy. *Journal of Aircraft*. 37(4). July-August. Also, see Kostoff, R. N., Green, K. A., Toothman, D. R., and Humenik, J. A. *Database tomography applied to an aircraft science and technology investment strategy*. TR NAWCAD PAX/RTR-2000/84, Naval Air Warfare Center. Aircraft Division. Patuxent River, MD.

Kostoff, R. N. (2000c). The underpublishing of science and technology results. *The Scientist*. 1 May.

Kostoff, RN, DeMarco RA. (2001a). Science and technology text mining. *Analytical Chemistry*. 73:13. 370-378A.

Kostoff RN, Del Rio JA, García, EO, Ramírez AM, Humenik JA. (2001b). Citation mining: integrating text mining and bibliometrics for research user profiling. *Journal of the American Society for Information Science and Technology*. 52:13. 1148-1156.

Kostoff, R. N., Miller, R., Tshiteya, R. (2001c). Advanced technology development peer review - A case study. *R&D Management*. July.

Kostoff, RN, Hartley J. (2001d). Structured abstracts for technical journals. *Science*. 5519:1067a. 292.

Kostoff, R. N., Tshiteya, R., Pfeil, K. M., and Humenik, J. A. (2002a) Electrochemical Power Source Roadmaps using Bibliometrics and Database Tomography. *Journal of Power Sources*. 110:1. 163-176.

Kostoff, RN. (2002b). Stimulating innovation. *International Handbook of Innovation*. In Press.

Mansfield, E. (1980). Basic research and productivity increase in manufacturing. *The American Economic Review*. 70(5). December.

Mansfield, E. (1991). Academic research and industrial innovation. *Research Policy*. 20.

Metzger, N., Zare, R. N. (1999). Interdisciplinary research: from belief to reality. *Science*. 283. 29 January. 642-643.

Naiman, R. J. (1999). A perspective on interdisciplinary science. *Ecosystems*. 2. 292-295.

Patton, J.D. (2002). The role of problem pioneers in creative innovation. *Creativity Research Journal*. 14(1). 111-126.

Siau, K. (1996). Electronic creativity techniques for organizational innovation. *Journal Of Creative Behavior*. 30(4). 283-293.

Smalheiser, N.R., Swanson, D.R. (1994). Assessing a gap in the biomedical literature – magnesium – deficiency and neurologic disease, *Neuroscience Research Communications*. 15(1).

Smalheiser, N.R., Swanson, D.R. (1998a). Calcium-independent phospholipase a (2) and schizophrenia. *Archives General Psychiatry*. 55(8).

Smalheiser, N.R., Swanson, D.R. (1998b). Using ARROWSMITH: a computer assisted approach to formulating and assessing scientific hypotheses. *Computational Methods and Progress in Biology*. 57(3).

Spender, J.C., Grant, R.M. (1996). (Eds.). *Knowledge and the firm*. Special Issue, Strategic Management Journal. 17(5-9).

Swanson, D.R. (1986). Fish oil, Raynauds syndrome, and undiscovered public knowledge. *Perspectives in Bioogy and Medicine*. 30(1). 7-18.

Swanson, D.R., Smalheiser, N.R. (1997). An interactive system for finding complementary literatures: a stimulus to scientific discovery. *Artificial Intelligence.* 91(2). 183-203.

Swanson, D.R. (1999). Computer – assisted search for novel implicit connections in text databases. *Abstracts of Papers of the American Chemical Society.* 217.

Swanson DR, Smalheiser NR, Bookstein A. 2001. Information discovery from complementary literatures: categorizing viruses as potential weapons. *Journal of the American Society for Information Science and Technology.* 52: 10. 797-812.

Taggar, S. (2001). Group composition, creative synergy, and group performance. *Journal Of Creative Behavior.* 35(4). 261-286.

Terleckyj, N. (1985) Measuring economic effects of federal R&D expenditures: recent history with special emphasis on federal R&D performed in industry. Presented at NAS Workshop on *'The Federal Role in Research and Development'.* November.

Terleckyj, N. (1977). *State of science and research: some new indicators.* Westview Press. Boulder, CO.

Van de Klundert, T.C.M.J., Palm, F.C. (1998). (Eds.) *Market dynamics and innovation.* Special issue, Economist. 146(3). 387-390. October.

Weeber M, Klein H, de Jong-van den Berg LTW, Vos R. (2001). Using concepts in literature-based discovery: simulating Swanson's Raynaud-fish oil and migraine-magnesium discoveries. *Journal of the American Society for Information Science and Technology.* 52:7. 548-557.

Wenger, W. (1999). *Discovering the obvious: techniques of original, inspired scientific discovery, technical invention, and innovation.* Psychegenics Press. P. O. Box 332. Gaithersburg, Maryland.

APPENDIX 1 - LITERATURE APPROACH

A. Overview

The theoretical basis of the literature approach mirrors the scientific process in many ways. Information from diverse literatures, with relevant interfaces, is

examined. All information is first analyzed and then synthesized to produce discovery and innovation. Initial work (Swanson, 1986; Gordon, 1996) examined three variable classes or themes (c, b, a) in two literature categories (C and B) using two different approaches (start with "c," determine "b," then determine "a; " start with "c" and "a," then determine "b").

The principal thematic variables determine a thematic literature. From the previous example, if Raynaud's disease is the thematic variable specified initially, then the corresponding thematic literature might be all the papers in a given database that contain the phrase Raynaud's disease. The remaining thematic variables and literatures are determined by applying different algorithms to the initial thematic literature and subsequent derived literatures. Again, from the previous example, an algorithm would be applied to the Raynaud's disease thematic literature to determine the thematic variable blood viscosity, and a derived literature could then be determined as all the papers in a given database that contain the phrase 'blood viscosity'.

The first approach in the initial reported work (Swanson, 1986; Gordon, 1996) could be viewed as addressing the question: What variables "a" could influence variable "c" through mechanisms "b", or, in the example described above, "What treatment factors "a" could influence Raynaud's disease "c" through the different mechanisms "b." This approach started with thematic variable "c" (e.g., Raynaud's disease), and used this variable to develop thematic literature C. Algorithms were applied to this thematic literature database to identify thematic variable "b" values (b1, b2, etc., representing characteristics such as blood viscosity, blood flow, blood platelets, poor circulation, and others) closely linked to thematic variable "c." Each value or theme of variable "b" (b1, b2, etc.) was used to develop a thematic literature B1, B2, etc. Algorithms were applied to each of the thematic B literatures to identify thematic variable "a" values (a1, a2, etc. representing characteristics such as fish oil, eicosapentaenoic acid, and others) closely linked to the specific thematic variable "b" of each thematic B literature. Values of the thematic "a" variables in each of the thematic B literatures not found in thematic literature C defined a subset of the thematic B literatures that was disjoint from thematic literature C (e.g., the term "fish oil" was not found in the Raynaud's disease literature). These disjoint thematic "a" variables and their associated thematic B literature subsets became candidates for discovery and innovation.

The other approach reported could be viewed as addressing the question: What are the mechanisms "b" through which variable "a" could impact variable "c." This approach started with variables "c" and "a", and their associated literatures C and

A, and identified variables "b" that were linked to both variables "c" and "a". The same types of algorithms as in the first approach were used to identify closely linked variables, and the requirement for disjointness between literatures C and A was used as a basis for discovery.

From the experience of these two approaches, it becomes clear that the independent and dependent variables chosen, and the algorithmic approach selected, depend on the question being asked. Further examination shows that other approaches beyond these two are possible to answer other questions. The present chapter examines seven approaches to generate innovation and discovery that are structured to answer seven different questions, and shows how the algorithms and techniques developed in Database Tomography are used in these approaches.

B. Specific Approaches

The following discussion will be limited to scenarios of three variables "a", "b", "c", and two literatures. In future studies, more complex cases could be candidates for analysis and experimentation.

For the simple two literature/ three variable case, seven separate generic cases are possible, where the variables specified can be viewed as "independent" and the variables determined can be viewed as "dependent:"

(1) specify "a," determine "b" and "c"; (2) specify "c," determine "a" and "b";
(3) specify "b," determine "a" and "c"; (4) specify "a" and "c," determine "b";
(5) specify "a" and "b," determine "c"; (6) specify "b" and "c," determine "a";
(7) specify "a" and "b" and "c," validate linkage existence.

Cases (1), (2), and (3) are the most open-ended and least constrained. In each case, one variable is specified, and the other two are determined using the DT algorithms, the condition of disjointness and, most importantly, expert judgement. Cases (4), (5), and (6) are more constrained, since two variables are specified, and the third is determined using similar processes to the above. Case (7) is fully constrained, and its purpose is to ascertain literature support for validation of a hypothetical relation between specified values of the three variables. Cases (4) and (5) are subsets of case (1); cases (4) and (6) are subsets of case (2); cases (5) and (6) are subsets of case (3); Case (7) is a subset of cases (1) through (6). The solution mechanics for each of these seven cases will now be outlined.

1. Opportunity Driven

This first case addresses the question, "What are the potential variable 'c' impacts that could result from variable 'a,' and what are the variable 'b' mechanisms through which these impacts occur?" One specific variant of this question is of particular interest and importance to the science and technology community, "What are the potential impacts on research, development, systems, and operations that could result from research on a given topic?"

If the generic question of this first case is applied to the above example for the case where variable "a" is "fish oil" only, it could be phrased as, "What are the potential impacts or benefits (positive or negative) resulting from fish oil that would not be obvious from examining the fish oil literature alone?" This is an open-ended question, and places no restrictions on the mechanisms "b" or the types of impact "c." The first case is represented schematically as:

a----->b----->c.

Here, "a" is the independent variable, and "b" and "c" are the dependent variables that result from the solution process. The operational sequence is to start with the variable "a" and generate a literature A. Again following the above example and using the abbreviations FO (fish oil), BV (blood viscosity), and RD (Raynaud's disease), this means that the process would start by identifying the FO literature (call this A1). Many approaches could be used to define this literature; the approach recommended here is the one used in recent Database Tomography studies (Kostoff, 2000a, 2000b) for defining literatures. As an example of one literature definition approach, the iterative Simulated Nucleation method (Kostoff, 1997b) would be used to identify all the papers in the Science Citation Index which contained FO (and other related terms in the query) in the title, keywords, and abstract fields. This collection of papers would constitute the FO literature

The next step in the process is be to identify the variables "b" (b1, b2, ...) linked closely to variable "a1," and then identify the literatures B associated with variable "b" (B1, B2, ... the BV literatures). For this step, the proximity analysis method used in the recent Database Tomography studies (or other co-occurrence techniques) would be employed. For a journal-based database, this method conceptually identifies phrases in paper titles or abstracts or main texts physically located near the term of interest. As an example, if the term of interest in a given database is Raynaud's disease, then the proximity analysis method would provide a list of all phrases in close physical proximity to the term Raynaud's disease for all

occurrences of this term in the text. The proximity analysis approach of Database Tomography is based on the experimental findings that phrases within a semantic boundary (same sentence, paragraph, etc.) located physically close to the term of interest are contextually and conceptually close to the term of interest. Continuing the above example, this step uses the proximity analysis of Database Tomography to identify phrases in the FO literature physically close to the term FO, such as "b1," "b2," etc.

For each of these identified phrases "b1," "b2," etc. , a literature (B1, B2, ...) is established by querying the SCI. The next step is, for each of these B literatures, to identify the linked variables "c" (c1, c2, ...) The process used to identify the variables "b1," "b2," etc. linked to variable "a1" is repeated to obtain the variables "c1," "c2," etc. linked to each value of variable "b." The subsets of the B literatures which are disjoint from literature A1 (e.g., the B literatures which don't contain the term FO) must then be identified, and the variables "c" (and their associated linking mechanisms "b" to variable "a1") within these disjoint B literature subsets then become the candidates for discovery and innovation.

It is obvious that the process can easily mushroom out of control unless stringent limiting constraints are placed on the number of B literatures and "c" variables selected. For example, suppose that three "b" variables "b1," "b2," "b3" (and their associated three B literatures (B1, B2, B3) are identified as closely linked to FO. Suppose also that each of these three "b" variables is closely linked to five "c" variables. Then four literature searches are required (A1, B1, B2, B3), and fifteen abc linked pathways must be examined for disjointness and discovery, according to the following:

a1--->b1--->c11; a1--->b1--->c12; a1--->b1--->c13; a1--->b1--->c14; a1--->b1--->c15;
a1--->b2--->c21; a1--->b2--->c22; a1--->b2--->c23; a1--->b2--->c24; a1--->b2--->c25;
a1--->b3--->c31; a1--->b3--->c32; a1--->b3--->c33; a1--->b3--->c34; a1--->b3--->c35

In reality, there will be hundreds, if not thousands, of candidate "b" and "c" variables. However, there are different ways by which the "b" and "c" variables can be sharply limited in number. First, the analysts performing the study would eliminate all non-technical content phrases that passed through the trivial word filter in the Database Tomography algorithm. Second, the numerical indices for each phrase generated by the Database Tomography proximity algorithm would be

used as one figure of merit for pre-selection of key phrases. Third, those "c" variables that reappear in different abc pathways would have a higher priority for selection. Fourth, analyst judgement would be applied to weight the potential value of the different abc pathways in computing figures of merit.

The literature searches and proximity analyses are fairly straightforward, and have been refined in the Database Tomography process. The main intellectual efforts must be focused on prioritizing and reducing the number of linked variables or literatures to be examined, and interpreting the relationships among the final disjoint literatures to generate potential discovery relationships.

2. Requirements Driven

This second case addresses the question, "What are the variables 'a' that could impact variable 'c,' and what are the variable 'b' mechanisms by which these impacts are produced?" Applied to the above example for the case where "c" is Raynaud's disease only, it could be phrased as "What are the factors and their associated mechanisms that could impact the course of Raynaud's disease that would not be obvious from examining the Raynaud's disease literature alone?" This second case is represented schematically as:

a<-----b<-----c

Here, "c" is the independent variable, and "b" and "a" become the dependent variables. The operational sequence is to start with variable "c," and generate a literature C. Again following the above example, this means that the process would start by identifying the RD literature (call this C1). The same literature definition process as in the first case would be used. The next step would be to identify the linked variables "b" (b1, b2, etc.) to variable "c1," and then their associated literatures B (B1, B2, the BV literatures). For this step, the proximity analysis method used in the recent DT studies would be employed again as in the first case. Continuing the above example, this step uses the proximity analysis of DT to identify phrases in the RD literature physically close to the term RD, such as "b1," "b2," etc.

For each of these identified phrases b1, b2, etc. a literature (B1, B2, etc.) is established by querying the SCI. The next step is, for each of these B literatures, to identify the variables "a" (a1, a2, etc.) linked to variable "b." The process used to identify the variables "b1," "b2," etc. linked to variable "c1" is repeated to obtain the variables "a1," "a2," etc. linked to each value of variable "b." The subsets of

the B literatures that are disjoint from literature C1 (e.g., the B literatures which don't contain the term RD) must then be identified, and the variables "a" within these disjoint B literature subsets (and their associated linking mechanisms "b" to variable "c1") then become candidates for discovery and subsequent innovation. The same stringent limits on variables and literatures used in the first case are applicable here.

3. Mechanism Driven

The third case addresses the question, "For a given mechanism 'b,' what are the variables 'a' that could impact the variables 'c'?" Applied to the above example for the case where "b" is blood viscosity, it could be phrased as, "What combinations of variables that could effect a change in the blood viscosity mechanism and could be impacted by a change in the blood viscosity mechanism are candidates for discovery that were not obvious from examining only the blood viscosity literature?" The third case is represented schematically as:

a<-----b----->c

Here, "b" is the independent variable, and "a" and "c" are dependent variables. The operational sequence starts with variable "b," and generates a literature B. Again following the above example, this means that the process would start by identifying and generating the BV literature (call this B1). The same literature definition and generation process as in the first case would be used. The next step would be to identify the variables "a" (a1, a2, etc.) and "c" (c1, c2, etc.) linked to variable "b1," and then their associated literatures A (A1, A2, the FO literatures) and C (C1, C2, the RD literatures). For this step, the proximity analysis method used in the first two cases would be employed for the BV literature (B1). Continuing the above example, this step uses the proximity analysis of DT to identify phrases in the BV literature physically close to the term BV, such as "a1," "a2," etc. (FO literature) and "c1," "c2," etc. (RD literature). However, an arbitrary step is required at this point, since the proximity analysis only provides the aggregate of the linked variables "a" and "c." The analyst is required to divide the aggregate linked variables obtained from the proximity analysis into two groups, "a" variables and "c" variables. In the above example, the proximity analysis would generate the linked variables such as fish oil and Raynaud's disease. The analyst would be required to specify two categorizations for these variables, such as "dietary factors" for the "a" variables and "diseases" for the "c" variables. This step will depend heavily on the analyst's expertise in the technical area and ability to create taxonomies.

The next step is to identify/ generate A and C literatures using the approach described above. The final step is to identify the subsets of A literatures and C literatures that are disjoint. Each group of articles from the A literature and the C literature that contains a "b1" variable is considered to be a linked group. The subsets of these literatures that are linked through the common "b1" variable and that are disjoint (i.e., the C literature does not contain the "a" variable and the A literature does not contain the "c" variable) must then be identified. The variables "a" and "c" within these disjoint A and C literature subsets linked through the "b1" variable then become the candidates for discovery and subsequent innovation. The same stringent limits on variables and literatures used in the first approach are applicable here.

4. Opportunity-Requirements Driven

This fourth case addresses the question, "What are the mechanisms 'b' through which variable 'a' could impact variable 'c'?" Applied to the above example for the case where "c" is Raynaud's disease only, and "a" is fish oil only, it could be phrased as, "What are the mechanisms through which fish oil could impact Raynaud's disease that would not be obvious from examining only the Raynaud's disease literature or the fish oil literature?" The fourth case is represented schematically as:

a----->b<-----c

Here, variables "a" and "c" are independent, and variable "b" is the dependent variable. The operational sequence is to start with the variable "c," and generate a literature C, and with variable "a," and generate a literature A. Again following the above example, this means that the process would start by generating the RD literature (call this C1) and the FO literature (call this A1). The same literature definition and generation process as in the first case would be used. The next step would be to identify the linked variables "b," and then their associated literatures B for both the A1 literature and the C1literature. For this step, the proximity analysis method used in the first two approaches would be employed, for the FO literature (A1) and the RD literature (C1). Continuing the above example, this step uses the proximity analysis of DT to identify phrases in the RD literature physically close to the term RD, such as "b1," "b2," etc. and to identify phrases in the FO literature physically close to the term FO, such as b51, b52, etc. The next step is to identify the subsets of the A1 literature and C1 literature that are linked. Each group of articles from the A1literature and the C1 literature that contains a

"b" variable is considered to be a linked group. The subsets of these literatures linked through the common "b" variables that are disjoint (i.e., the C1 sub-literature that does not contain the "a1" variable and the A1 sub-literature that does not contain the 'c1' variable) must then be identified, and the variables "b" within these disjoint A1 and C1 literature subsets then become the candidates for discovery and subsequent innovation. The same stringent limits on variables and literatures used in the first case are applicable here.

5. Opportunity-Mechanism Driven

The fifth case addresses the question, "What are the variables 'c' which could be impacted by variable 'a' through mechanism(s) 'b'?" While the schematic shown for this case is identical to that of case 1, the two schematics should be interpreted differently. In case 1, the intermediate mechanism(s) "b" are not specified beforehand, but are a result of the solution process. In the present case, these "b" mechanism(s) are specified beforehand. Applied to the above example for the case where "b" is blood viscosity only, and "a" is fish oil only, the question in this case could be phrased as, "What abnormalities could be influenced from the impact of fish oil on blood viscosity that would not be obvious from examining only the abnormality's literature or the fish oil literature?" The fifth case is represented schematically as:

a----->b----->c

Here, "a" and "b" are the independent variables, and "c" is the dependent variable. The operational sequence is to start with the variable "a," and generate a literature A, and with variable "b," generate a literature B. Again following the above example, this means that the process would start by generating the FO literature (A1) and the BV literature (B1). The same literature definition and generation process as in the first case would be used. The next step would be to identify the linked variables "c," and then their associated literatures C (the collection of RD literatures) for the B1 literature. For this step, the proximity analysis method used in the previous cases would be employed for the B1 literature only. Continuing as before, this step uses the proximity analysis of DT to identify phrases in the BV literature physically close to the term BV, such as "c1," "c2," etc. The resulting C literatures are automatically linked to the A1 literature through the linking variable "b1." The "c" variables which are disjoint to the A1 literature (i.e., the C sub-literature that does not contain the "a1" variable and the A1 literature does not contain the "c" variables) must be identified, and become the candidates for

discovery and subsequent innovation. The same stringent limits on variables and literatures used in the first case are applicable here.

6. Requirements-Mechanism Driven

The sixth case addresses the question, "What are the variables 'a' that could impact variable 'c' through mechanism 'b'?" Applied to the above example for the case where "b" is blood viscosity only, and "a" is fish oil only, it could be phrased as, "What factors could impact Raynaud's disease by impacting blood viscosity that would not be obvious from examining only the factors' literature or the Raynaud's disease literature?" The sixth approach is represented schematically as:

a<-----b<-----c

Here, "b" and "c" are the independent variables, and "a" is the dependent variable. The operational sequence is to start with the variable "c," and generate a literature C, and with variable "b," and generate a literature B. Again, this means that the process would start by identifying and generating the RD literature (C1) and the BV literature (B1). The same literature definition and generation process as in the first case would be used. The next step would be to identify the linked row of variables "a" (a1, a2, etc.), and then their associated literatures A (the FO literatures) for the B1 literature. For this step, the proximity analysis method used in the previous cases would be employed, for the B1 literature only. Continuing as before, this step uses the proximity analysis of DT to identify phrases in the BV literature physically close to the term BV, such as "a1," "a2," etc. The resulting A literatures are automatically linked to the C1 literature through the linking variable "b1." The "a" variables which are disjoint to the C1 literature (i.e., the A sub-literature does not contain the "c1" variable and the C1 literature does not contain the "a" variables) must be identified, and become the candidates for discovery and subsequent innovation. The same stringent limits on variables and literatures used in the first case are applicable here.

7. Opportunity-Mechanism-Requirements Validation

The seventh case addresses the question, "Does the literature support the possibility that variable 'a' could impact variable 'c' through mechanism 'b'?" Applied to the above example for the case where "a" is fish oil only, "b" is blood viscosity only, and "c" is Raynaud's disease only, it could be phrased as, "Does the literature support the possibility that fish oil could impact Raynaud's Disease by altering blood viscosity in a way that would not be obvious from examining only

the fish oil literature or the Raynaud's disease literature?" The seventh approach is represented schematically as:

a<----->b<----->c

Here, "a" and "b" and "c" are independent variables. The operational sequence could start with either "a" or "b" or "c." For the present discussion, the operational sequence starts with the variable "b," and generates literature B. Again following the above example, this means that the process would start by identifying and generating the BV literature (B1). The same literature generation process as in the first approach would be used. The next step would be to extract the B1 sub-literatures which contain the variables "a1" (literature A1) and "c1" (literature C1).

The final step is to validate the existence of disjoint A1 and C1 sub-literatures (i.e., A1 sub-literature that does not contain the "c1" variable and a C1 literature that does not contain the "a1" variable). The "a1"-"b1"-"c1" sequence then becomes a candidate for discovery and subsequent innovation. The same stringent limits on variables and literatures used in the first approach are applicable here.

APPENDIX 2: CROSSING THE BRIDGE: INTERDISCIPLINARY WORKSHOPS FOR INNOVATION.

BACKGROUND

The Office of Naval Research established a series of workshops in 1997 aimed at promoting innovation while also enhancing organization, category, and discipline diversity components. The focus of the first novel workshop founded on this plan was "Autonomous Flying Systems," an area of perceived long-term interest to not only the Navy and Department of Defense, but also to the National Aeronautics and Space Administration and other governmental and industrial organizations. The process employed was designed starting with a clean slate and was intended for application to very significant technical challenges. The present appendix further describes the process that was used to identify the technical theme of the workshop, select the participants, and conduct all three phases of the total workshop.

WORKSHOP THEME IDENTIFICATION

It was decided that the initial workshop theme should 1) focus on problems related to the main science and technology emphasis area of the author's home

organization, Strike Technology, and 2) help establish the most supportive environment for innovation. The problem selected should be focused and understandable, and it should have a generic technical base amenable to soliciting people from many different disciplines. The topic finally selected was autonomous control of unmanned air vehicles, including takeoff and landing from limited areas on smaller Navy ships. It was apparent that the underlying science and technology permeated many different disciplines, including aerodynamics, controls, structures, communications, guidance, navigation, propulsion, sensing, and systems integration. Also, the naval applications for some aspects of this problem were sufficiently unique that probably not a great deal of work had been done in this area. Subsequent literature analyses validated this assumption.

Present naval air systems are either manned (most aircraft) or tele-operated, semi-autonomous (weapons and some aircraft). The weapons are a mix ranging from "dumb" bombs and shells to "smart" missiles. The future trend is toward "smart" autonomous or semiautonomous aircraft and weapons. Since a major role of the Office of Naval Research is to proactively address the technology that will influence future naval forces, it seemed natural to examine science and technology roadblocks on the path to unmanned autonomous "smart" flight systems. Consequently, the focus of the initial workshop was defined as identification of the fundamental operational principles of autonomous flying systems over a fairly wide range of flight environments. In particular, the workshop was aimed at examining what had been learned about autonomous or semiautonomous operation from the animal (mainly flying) kingdom and from other unmanned autonomous/ semiautonomous tele-operated systems such as autonomous underwater vehicles and locomoted robots. Animals are now being studied as integrated systems by scientists on the forefront of biological research. The issues of aerodynamics, flight mechanics, dynamic reconfiguration, materials, control, neuro-sciences, and locomotion are not being studied as separate disciplines by these scientists, but rather are being studied in parallel in the same animal system and in their relation to the function and mission of the animal system. While this integrative biological research is in its infancy, and results are only starting to emerge, the time seemed appropriate for assembling these diverse groups and exploiting their synergy. Not only could there be benefit to the Navy from such cross-discipline interaction, but benefit could be possible for each of the contributing disciplines as well.

A major thrust of the workshop was projected to be identification of the autonomous operational principles for each unique system and the relation of these principles to mission and function, then extraction of the generic operational principles that underlay all the systems, both biological and man-made. It was

hoped that the cross fertilization of disciplines would be able to further elucidate and clarify the more important generic concepts, and then provide insight that could be utilized to enhance the autonomous operation of naval flying systems.

PARTICIPANT SELECTION

Once the theme of the workshop was established, a sub-theme taxonomy was developed to focus the agenda and to identify workshop participants. A dual approach was followed to generate the taxonomy.

Discussions were held with agency experts on the generic theme concerning the taxonomy structure. In parallel, the Science Citation Index was queried for papers related to the generic theme. Both bibliometric and computational linguistics analyses of these papers were performed to provide strategic maps of the topical area, identifying key performers, journals, institutions, and their relations to the technical themes and sub-themes of the workshop. A taxonomy was constructed based on these strategic maps. (For a description of how the bibliometric and computational analyses are combined to generate strategic maps, see Kostoff (1998, 1999b, 2000a, 2000b, 2002a)).

Both of these taxonomy sources, in-house experts and the Science Citation Index, then provided initial candidates for participation in the workshop. These candidates were contacted, and asked to suggest additional candidates. This procedure continued until a large pool of potential candidates was established. Three main selection criteria for workshop participants were established;

(1) Multiple recommendations,
(2) Significant publications is the field, and
(3) Literature citations.

These three criteria were tempered with judgement to insure that bright young individuals, who had not yet established a track record, were not excluded from the pool, and that the panel as a whole had the correct level of discipline, category, and organization balance. In addition, a guideline was established that all workshop attendees would be active participants, so the number of attendees was limited to facilitate discussion and interactions.

All these constraints, guidelines, and selection criteria were used to arrive at the final panel size and structure. The result was a panel of slightly more than twenty people representing a mix of disciplines that included biologists (experts in bird,

bat, frog, fish, or insect studies), robotics, artificial intelligence, controls, autonomous aircraft, fluid dynamics, sensors, neuroscience, cognitive science, autonomous underwater vehicles, aerodynamics, propulsion, and avionics.

OVERVIEW OF WORKSHOP PROCESS STEPS

(1) Workshop Buildup

The buildup period for the workshop in question started about two months before the meeting. Specific guidance for the conduct of the workshop was sent to the participants by e-mail, including a statement of the naval technical problems to be addressed. The technical component of the buildup phase was then conducted by e-mail.

The main purpose of this buildup phase technical component was to have each participant generate new ideas from his/ her discipline for all other participants to consider. The other participants could then dialogue by e-mail to clarify/ modify/ embellish these ideas. At a minimum, even if no dialogue resulted, there would be a gestation period of about two months for each participant to absorb these concepts from other disciplines. Specifically, each participant was requested to:

- Submit a half dozen leading edge capabilities or accomplishments in his/her discipline(s) that could potentially impact the naval technical problems; and

- Identify several leading edge capabilities or accomplishments projected in his/her discipline(s) over the next decade that could potentially influence the naval technical problems; and

Submit a few leading edge capabilities or accomplishments in his/ her discipline(s) whose impact on the naval technical problems was not obvious to him/ her, but might be obvious to someone else.

The participants were free to comment on potential relations among any of the capabilities, accomplishments, or combinations of capabilities and accomplishments, and any of the naval technical problems, or combinations of problems. All of the comments received were then sent to all the participants. This exercise helped stimulate the thinking of the participants, and provided a documented record of the process. One of the functions of the participants from the author's organization was to facilitate and stimulate dialogue by raising questions and issues on the submitted information.

If any of the participants saw a capability or accomplishment from another participant that could impact a problem in his/her discipline, but not impact a naval technical problem, then the two participants were free to dialogue together without informing all the participants. However, these two participants engaged in independent dialogue were requested to keep a record of their exchange that might be included with the final workshop report as potential discovery. This would cover the real possibility of discovery occurring in topics other than the one targeted.

(2) Workshop Meeting

As a result of the ideas presented during the buildup phase, it appeared that the seeds existed for a new science and technology program on Autonomous Flying Systems. Therefore, an agenda was sent to the participants with further guidance to address promising science and technology opportunities at the workshop, that would serve as the foundation of such a program. Specifically, the participants were asked to address the following issues at the workshop:

- What are the present leading-edge capabilities in your discipline?

- What are the desired future capabilities in your discipline?

- What are the leading research opportunities in your discipline and what additional capabilities could they provide if successful?

- What is the level of risk of these opportunities successfully achieving their targets?

- How would these potentially enhanced capabilities contribute to, or translate into, improved understanding and/or operation of autonomous flying systems?

The meeting occurred on 10-11 December 1997 at ONR. Since some of the leading edge capabilities and potential accomplishments appeared to have applicability to naval technical problems (identified during the e-mail buildup period), the proponent for the capability or accomplishment item took the lead in fleshing out his/her ideas and leading the discussion at the meeting. As a result, the workshop meeting tended to evolve into full panel discussions on each of these potential capabilities.

There were two rounds of discussion at the workshop. The first round consisted of presentations and discussions by each proponent. The second round of the workshop consisted of each participant identifying his/her leading promising research opportunities.

(3) Workshop Cleanup

The participants were requested to provide any additional narrative information that added to or modified their ideas as a result of the workshop experience. The outcomes of the workshop included both the tangible and intangible.

Three immediate tangible outcomes were projected:

(1) A concept proposal for a science and technology program focused on Autonomous Flying Systems would be generated;

(2) Technical papers may be submitted to leading science journals based on innovations identified; and

(3) One or more papers on the complete workshop experience might be submitted to leading science journals.

In addition to developing specific topics, it was anticipated that new, un-exploited ideas in interdisciplinary research and development might surface during contact between panelists. These novel subjects might form the basis of additional workshops. In addition, extensive lessons were learned as a result of the workshop process. These lessons were summarized in section II-B.

APPENDIX 3–MULTIPLE DISCIPLINE SELECTION FOR INNOVATION

Overview

This appendix describes the use of modern information technology to identify the balance of research disciplines required to enhance innovation, including when multi-disciplinary or inter-disciplinary approaches should be used.

Introduction

In complex research problems, addressing only one or a few of the component disciplines may result in fragmented or perhaps misleading results due to neglect of discipline inter-dependencies. However, even if the many disciplinary facets of a complex research problem are addressed, the method of integration of the multiple facets can impact the solution of the problem. Research that includes multiple disciplines but maintains their distinctiveness is multi-disciplinary (Collins 2002). Such research may not include joint planning, management, and review of the multiple disciplines. Research that integrates the multiple disciplines to effectively form a new unified discipline is inter-disciplinary. Even if all of the multiple component disciplines are addressed separately in a multi-disciplinary approach, the final research product will not have the same quality as a unified research product resulting from an inter-disciplinary study, especially if the different disciplines impact each other strongly.

Another strong motivation for examining multiple disciplines is increased evidence that there are common underlying themes across many research fields. For example, the same equations are used to model phenomena in some very diverse disciplines, such as the modeling of chaotic behavior. Appropriate inter-discipline research and information transfer can allow findings and insights from one discipline to be extrapolated and exploited by another, perhaps very disparate, discipline.

Paradoxically, in parallel with the increasing need for inter-disciplinary projects, researchers have become much more specialized by necessity. The massive global expansion of technical literatures and other science and technology products reduces the time available for researchers to remain current in their own specialty disciplines, much less to become familiar with progress in other disciplines. In addition to lack of time, they also have many other dis-incentives to participate in inter-disciplinary projects (see Appendix 3A). If there are no external incentives offered for inter-disciplinary research, most researchers will take the path of least resistance, and restrict their research projects within their own, or very closely related, disciplines.

In recent years, research sponsoring agencies have decided there is merit to inter-disciplinary research, and have provided incentives for the proposal and establishment of such programs. In many cases, the result has been programs that are inter-disciplinary on paper only. They are not managed or reviewed as a cohesive inter-disciplinary unit, but are managed and reviewed (in practice) as fragmented separate programs. In other cases, programs (and facilities) have been advertised as inter-disciplinary when in reality each 'discipline' is a minor variant

of a single discipline (e.g., Physics/ Materials, where the materials group members are basically physicists who happen to be focusing on the physics of materials). The number of true inter-disciplinary projects and programs that incorporate distinctly different disciplines, but are selected, managed, reviewed, and transitioned as cohesive units, is a small percentage of all research conducted.

Further, it is difficult to objectively gauge the effectiveness of these inter-disciplinary efforts. The metrics used for these assessments, such as numbers of paper authors from different disciplines or mixes of discipline funding under program managers, are very incomplete. These quantitative metrics are amenable to manipulation, can be deceptive, and intrinsically do not describe the quality of the discipline mixing process. Most egregiously, they do not separate artificial inter-disciplinary projects, such as the Physics/ Materials example above, from coherent projects consisting of relatively disparate disciplines.

However, it is not necessary to conduct all research programs as inter-disciplinary. There are some tangible and intangible costs involved in conducting inter-disciplinary programs, due to the overhead required to integrate diverse technical cultures and traditions (see Appendix 3A). A program should be conducted as inter-disciplinary only if a strong diverse mix of disciplines is required to fully address its research objectives. There is no intrinsic virtue to conducting projects or programs as inter-disciplinary, unless it can be demonstrated that they fundamentally require an inter-disciplinary approach for maximum advancement.

Process Concept

The fundamental thesis of this appendix is that the mix of disciplines used in the conduct of a science and technology program should correspond to the multiple discipline requirements of the program. A systematic three-step process (based on the use of modern information technology) is proposed for determining the relationship of the disciplines required to conduct a science and technology program to the disciplines selected. The first step in the process is *identification of the multiple disciplines* that could have some impact on the research problem. The second step is *determination of the cost-effectiveness* (importance versus costs) of employing all the disciplines that could potentially impact the problem. The third step is *provision of incentives/ mandates* to the performers for incorporating those required disciplines that will contribute to the problem's solution cost-effectively.

Process Mechanics

Background

The proposed three-step process is based on the literature-based discovery technique described in Appendix 1(Swanson 1986, Swanson and Smalheiser 1997, Gordon and Lindsay 1996, Weeber et al 2001, Kostoff 1999, Kostoff 2002b). In literature-based discovery, identification and merging of concepts from very different technical disciplines are not options; they are requirements.

The literature-based discovery studies that have been performed confirm the parochialism of researchers in the specific disciplines studied. Consider Swanson's initial paper on literature-based discovery (Swanson 1986), in which he hypothesized that Fish Oil/ Eicosapentaenoic Acid could alleviate some symptoms of Raynaud's Disease (later confirmed by laboratory and clinical tests). The Raynaud's Disease researchers were not aware (based on what could be deduced from the literature analysis) of the Fish Oil literature, and the Fish Oil researchers were not aware of the Raynaud's literature.

Further, a recent bio-terrorism-related literature-based discovery study (Swanson et al 2001) identified viruses that are not recognized today as bio-warfare agents, but have the characteristics to be modified into bio-warfare agents. Such viral agents pose a special threat, since their use would contain the element of surprise. For such agents, there would be no vaccines for prevention, no detection, and perhaps no therapies, and the potential destructive consequences would be far greater than those of the anthrax bacterium. These viruses had gone un-recognized as candidate bio-warfare agents by the technical specialty communities. The two main bio-warfare agent characteristics, virus pathogenicity and virus transmissibility, had been studied by two disjoint research communities that were not familiar with each other's literatures (based on what could be deduced from the literature analysis).

First and Second Steps

The first step in the process is to perform a literature-based discovery analysis of the research problem prior to initiation of a research project. The output would consist of identifying:

1) technical disciplines that could potentially contribute to advances in the research problem;
2) experts within these disciplines; and possibly (not necessarily)
3) potential problem solutions.

In the tandem second step, the proposers or principal investigators could then estimate the importance of each of the identified disciplines to the attainment of the research objectives, and use that as the basis for a strategy of constructing the research approach.

This second step would use the output from the literature-based discovery for convening the workshops or groups of experts described in the main body of the text, and in Appendix 2. The combination of literature-based discovery followed by guided workshops would eliminate the following deficiencies of standard workshops:

1) Small community representation
2) Parochialism; not all relevant disciplines represented
3) Human dynamics; can overwhelm technical discussions
4) High degree of subjectivity

as well as the following deficiencies of literature-based discovery:

1) Only a small fraction of R&D conducted gets published
2) Currency; there is a lag time in publication
3) Minimal human interaction for concept stimulation
4) A specific solution to the problem may not be identifiable from the literature alone

This combination would retain the strengths of each component to produce a systematic enhancement of the environment for stimulating innovation. In the workshop, the range of required disciplines would be clarified further, and disciplines added or subtracted to the proposed research approach as dictated by the additional costs and benefits to science and technology. In addition, if the literature-based discovery has generated discovery in the form of specific hypotheses to be tested, these could be discussed and sharpened further.

Third Step

The first two steps are mechanistic technology steps. They will work technically, although improvements in each are desirable and possible. The third step is the most difficult, since it involves incentives and the accompanying human issues of motivation, tradition, culture, and inertia. If progress is to be made in pursuing intrinsically inter-disciplinary research appropriately, mandates requiring at least

the first step of the hybrid process (literature-based discovery) are probably required initially. After the technical community becomes convinced of the benefits of incorporating literature-based discovery at the initiation of research projects, and becomes familiar with the process mechanics involved, then incentives can probably replace mandates for performing pre-project literature-based discovery.

There is precedent for these types of pre-project literature survey mandates. A number of Federal agencies require literature surveys before initiation of research projects. Since literature-based discovery (sans workshop) could be viewed as a sophisticated form of literature survey, introduction of a pre-project literature-based discovery requirement would in some sense be an extension of existing literature survey requirements.

SUMMARY AND CONCLUSIONS

A three-step process has been proposed for insuring selection of a comprehensive mix of research disciplines to address a research problem. The process is based on literature-based discovery to identify and select the comprehensive discipline mix before research is started. When appropriate, workshops can be convened using the information developed in the literature-based discovery component.

In this scenario, the literature-based discovery approach would serve as one block in the foundation of all research performed, in helping to objectively determine the mix of disciplines required to attain the research objectives. It may also provide discovery based on the literature studies alone. Even if actual discovery does not result from the literature phase alone, the fundamental value of literature-based discovery in determining discipline mixes for subsequent workshops and research program conduct remains un-diminished.

To insure that most of the potentially important disciplines are identified by the literature-based discovery process, more process development is required, and more variants of literature-based discovery are required. The quality and credibility of the literature-based discovery output depends on:

1) Study objectives; metrics used
2) Source databases used (e.g., Medline, Science Citation Index, Pascal)
3) Information retrieval techniques used
4) Record fields analyzed (e.g., Keywords, Titles, Abstracts, Full Text)

5) Analysis techniques, especially co-occurrence and clustering techniques (Kostoff et al, 2001a, 2001b, 2002a)
6) Most importantly, the people performing the analysis

Each variant of literature-based discovery will use one or more alternatives of these study components, and only very few literature-based discovery studies have been published so far. This expanded development of literature-based discovery has not yet been started.

This deficiency is particularly egregious relative to the present global threat from bio-terrorism. To the author's knowledge, Swanson et al (2001) was the only published literature-based discovery study to have addressed bio-warfare agent prediction. *One small study, using one approach, represents the total reported globalliterature-based discovery effort to prevent surprise by potential biowarfare agents that could be identified with publically available knowledge!* In what other area of science and technology is only one approach, no matter how good, used to solve a problem? Multiple literature-based discovery approaches, and multiple studies, are required to insure that as many candidate bio-warfare agents as possible are identified.

A national effort is needed to develop parallel literature-based discovery approaches, to insure that optimal methods are used to identify and integrate findings from disparate disciplines. Further, experiments are required to identify how the literature-based discovery results should be integrated with workshops to exploit these multi-disciplinary findings and maximize the potential for innovation. Finally, the requirement for incorporating literature-based discovery at the initiation of research projects, to insure that all relevant research reported and all potentially relevant disciplines are identified, should be mandated for all Federally-supported research. Such a process would identify research that required multiple disciplines for rapid advancement, as well as research that could produce acceptable results from mono-discipline analysis.

APPENDIX 3A – MULTI-DISCIPLINARY AND INTER-DISCIPLINARY RESEARCH BARRIERS

Some of the specific barriers to multi-disciplinary and inter-disciplinary research include Culture, Time, Evaluation, Publication, Employment, Funding, Promotion, and Recognition.

Culture

Different technical disciplines represent different cultures and traditions. Each culture has its own vocabulary, its own perspective on what constitutes evidence, its own standards of proof, its own definitions of truth, and its own traditions on how research is defined and performed. Merging of cultures and traditions for inter-disciplinary research requires communication, coordination, and consensus among cultures, and compromise from all parties. Additional time is required to structure inter-disciplinary proposals, and to plan the conduct of research projects (Bauer 1990, Naiman 1999).

Time

Inter-disciplinary research requires that each participant learn some aspects of the other participants' disciplines, including the cultures and traditions noted above. Time is required to learn these other technologies, cultures, traditions, and to effect the coordination and consensus processes. This time expenditure detracts from time spent on the mastery of a single discipline (Naiman 1999).

Evaluation

Peer review is the main and preferred type of research evaluation (Kostoff 1997). Traditionally, peer review has consisted mainly of judgements from mono-discipline reviewers, often in the same research area as the reviewee (Bruhn 1995, Metzger and Zare 1999, Butler 1998). Reviewers tend to give higher marks to in-depth advances made in a single discipline rather than less intense advances made across a wider range of disciplines.

Publication

Most ranked journals tend to have a strong mono-disciplinary mission, and many will even discourage submittal of broader-based inter-disciplinary manuscripts (Bruhn 1995, Butler 1998, Naiman 1999). The manuscript review process tends to have similar structure and reviewer parochialism problems for inter-disciplinary research as noted above under Evaluation. The document Abstract, the main vehicle for communicating research content across disciplines in the large databases such as Medline and Science Citation Index, is in many cases incomprehensible to all but the research area experts (Kostoff and Hartley 2001d).

Employment

Graduates with specialist degrees are often more marketable than generalists (Bruhn 1995). The problem lessens somewhat as employment in higher budget categories (transition to systems development) is pursued, due to natural merging of disciplines as focused technologies advance into broader systems.

Funding

Many of the large research-sponsoring organizations are structured along the lines of mono-discipline university departments. Their review panels tend to have similar structures, and have the same problems for multi/ inter-disciplinary research as noted above under Evaluation (Bruhn 1995, Butler 1998, Metzger and Zare 1999). In general, mono-discipline research proposals fare better than inter-disciplinary research proposals, except where programs have been specifically designed to fund inter-disciplinary research proposals.

Promotion

The reward system in universities is designed to recognize the research and scholarly contributions of individuals, not teams (Bruhn 1995, Metzger and Zare 1999). Tenure in universities is dependent on the number and quality of publications, and is helped by funds that researchers can attract. As shown above, publications and funding are easier to obtain in mono-disciplinary research, and therefore inter-disciplinary research is penalized further.

Recognition

National academies and other prestigious professional organizations and awards are almost wholly discipline-structured (Metzger and Zare 1999). Since recognition has some dependence on publications and citations, and in many cases on research empires established (funding obtained), mono-disciplinary advantages noted above for publications and funding flow into recognition as well.